Missing observations in the dynamic regression model

Nijman, T.E.; Palm, F.C.

Published in:
Econometrica

Publication date:
1984

Econometrica, Vol. 52, No. 6 (November, 1984)

MISSING OBSERVATIONS IN THE DYNAMIC REGRESSION MODEL

BY F. C. PALM AND TH. E. NIJMAN[1]

We consider the dynamic regression model with lagged endogenous variables and moving average disturbances, when some observations on the endogenous variable are missing. The available data are assumed to be sampled at regular intervals of length m and can be linear combinations of the realizations of the variable over a finite number of periods.

We discuss the identification of the parameters in the model. For some selected models, we evaluate the large sample variances of the maximum likelihood (ML) estimates for the incomplete data and complete data respectively. In this way, we get an indication of the loss of information when the data are incomplete.

Finally, we give some results for the effects on the properties of the OLS estimator, when interpolated series are substituted for the missing observations and we briefly discuss ways to obtain ML estimates. Our general conclusion is that when the sample is incomplete it is very important to use all available reliable a priori information to analyze the model.

1. INTRODUCTION

IN ECONOMETRIC ANALYSIS of time series, it is usually assumed that the relevant data consist of observations on the variables in the model pertaining to T subsequent time periods that are considered appropriate on a priori grounds. Attention has been drawn in the literature to the consequences of loosening these assumptions which will often not be met in applied work. One stream of contributions is concerned with the problem of missing observations. The problems of temporal aggregation in dynamic models form another related research topic that has received increasing attention in recent years.

In this paper we concentrate on the dynamic regression model with moving average disturbances when the endogenous variable is observed every mth period, as is usually the case for stock variables, or when only a linear aggregate for the m periods, such as a flow variable measured over the m periods, is observed. Formally, we assume that the endogenous variable y_t is generated by the following regression model:

$$\rho(L)y_t = \sum_{k=1}^{K} \beta_k x_{kt} + \theta(L)\varepsilon_t, \tag{1}$$

where $\rho(L) = 1 - \sum_{i=1}^{p} \rho_i L^i$, $\theta(L) = \sum_{j=0}^{q} \theta_j L^j$, L being the lag operator, and the ε_t's are independent normal variates with mean zero and variance σ^2, $\theta_0 = 1$, and the x_{kt}'s are strictly exogenous variables, i.e., x_{kt} is independent of ε_{t-j} for all t, j, and k. We assume that the standard conditions for identification of the parameters in (1) are satisfied. Furthermore, we assume that all x_{kt}'s are known ($k = 1, \ldots, K$; $t = 1, \ldots, T$), but that only linear combinations of the y_t's defined

[1] An earlier version of this paper was presented at the Econometric Society Meeting in Dublin, September 5–9, 1982. The authors thank *Econometrica*'s editors and a referee for their useful comments and R. J. Reichardt for his help in carrying out the computations.

by

$$\tilde{y}_t = \sum_{i=0}^{A} w_i y_{t-i}, \qquad \text{for } t \in T_m, \tag{2}$$

are observed, where T_m is defined as the set $T_m = \{m, 2m, \ldots, T\}$. The weights w_i are deterministic and known.

The problem of missing observations has been analyzed along two different lines. A simple formal way consists in deriving for instance quarterly data from yearly observations on the same series by minimizing some criterion function (see, e.g., Boot, Feibes, and Lisman [2]). A second approach consists in specifying a model in which the missing observations are explained by other variables. When exogenous variables are missing, one usually extends the model by introducing an equation that relates the unobserved exogenous variables to other explanatory variables. In this way, our results also apply to the case where observations on an exogenous variable are missing. Next, the unobserved realizations are integrated out (substituted for) to get the data generation process. The parameters of the model (1) for data and missing observations can be estimated, provided they are identified, and the model can be used to generate predictions of the missing observations.[2] Quite often, the parameters of the process for the observations will be subject to restrictions which follow from the specification of the joint process for observations and missing data. These restrictions allow for testing the validity of the initial model.

The problem of missing data in dynamic models has received attention in econometrics and time series analysis (see, e.g., Kmenta [9] and Dunsmuir [4] and the references cited therein). Usually, although not exclusively, estimation problems have been studied in the literature. This paper is focussed on the identification problem in relation with the sampling scheme in the dynamic regression model with missing endogenous variables. Results for univariate ARMA-models will also be discussed. Special attention will be paid to the loss of information due to incomplete sampling. Estimation problems will only be briefly considered.

The plan of the paper is as follows. In Section 2, we introduce some patterns for the data transformation (2) that are particularly relevant in economic applications. Section 3 is devoted to the identification of the parameters in the model. In Section 4, we examine the loss of information due to incomplete sampling by comparing the asymptotic efficiency of the maximum likelihood (ML) estimates for the model when data are missing with that of the estimates when all observations are available. In Section 5, we briefly consider the computation of ML

[2] When the missing data have to be predicted, identification of the parameters in (1) may be required. For example, for $m=2$, $K=1$, $p=1$, $g=0$, the conditional expectation of y_t given past observations on y_t and past and present observations on x_{1t} is $Ey_t = \rho y_{t-1} + \beta_1 x_{1t}$. When future observations on y_t and x_{1t} are also used, it becomes:

$$Ey_t = (1+\rho^2)^{-1}[\rho y_{t-1} + \rho y_{t+1} + \beta_1 x_{1t} - \beta_1 \rho x_{1t+1}].$$

Using the projection of y_t on the space for x_1 only yields predictions for y_t, $t \in T_2^c \to T_1 \backslash T_2$ or t being outside the sample period, which are less accurate than those using the observed values of y_t as well. When $\beta_1 = 0$, identification of ρ is required for one step ahead prediction of y_t.

estimates and present some results on the effect on parameter estimates in large samples of using interpolated data as proxies for the missing endogenous variables. Finally, in Section 6, some concluding remarks are given.

2. RELEVANT DATA TRANSFORMATIONS

In the previous section we assumed that the observations are in the form of the linear transformation (2) of the unknown data. Dropping the assumption of linearity would admittedly introduce new problems. Transformations of type (2) are somewhat restrictive however; several authors have discussed transformation patterns that do not fit into (2) (e.g., Dunsmuir and Robinson [5], Harvey and Pereira [7]). For economic time series, the most important cases that do not fit into (2) are perhaps the randomly missing observations and the (α, β)-sampling with $\alpha \neq 1$. In case of randomly missing observations the availability of an observation on y_t is determined by a probabilistic mechanism that is independent of the probability law according to which y_t is generated. The term (α, β)-sampling refers to a procedure in which the process under consideration is periodically observed for α consecutive periods and not observed for the next β consecutive periods. For the scheme in (2), $\alpha = 1$. Although the scheme (2) is restrictive, it is relevant in many economic applications. Moreover, some of the implications of the $(1, \beta)$-sampling remain valid for more general sampling schemes.

For a stock variable y_t, observations will often be available every mth period. If the data are generated by a quarterly model and observed on an annual basis, $m = 4$ and the coefficients in (2) are

$$A = 0, \quad w_0 = 1. \tag{3}$$

This set of coefficient values will be referred to as the skipped data pattern. If y_t is a flow variable, the total flow for m periods is usually observed, so that we have

$$A = m - 1, \quad w_i = 1 \qquad (i = 0, 1, 2, \ldots, m-1). \tag{4}$$

The scheme (2) is valid in other cases as well. Assume that the model (1) is formulated in first differences, that is $y_t = \Delta z_t = z_t - z_{t-1}$. If z_t is a stock variable observed every mth period $(t \in T_m)$ then $z_t - z_{t-m} = \sum_{i=0}^{m-1} y_{t-i} = \tilde{y}_t$, for $t = 2m, 3m, \ldots, T$, can be obtained. In this case, (4) applies as well. If z_t is a flow variable for which every mth sum of the last m realizations is observed, we have information on $\tilde{y}_t = (1 - L^m) \sum_{i=0}^{m-1} z_{t-i}$ which gives $\tilde{y}_t = (1 + L + \cdots + L^{m-1})^2 y_t$. These transformations of y_t, with y_t being generated by a static regression model, have been analyzed by Zellner and Montmarquette [16]. Similar patterns arise when the model (1) is formulated in second differences, $y_t = \Delta^2 z_t = z_t - 2z_{t-1} + z_{t-2}$, while we observe skipped data, z_t, $t \in T_2$ (assuming $m = 2$), so that

$$\tilde{y}_t = z_t - 2z_{t-2} + z_{t-4} = y_t + 2y_{t-1} + y_{t-2} \qquad (t = 6, 8, \ldots, T),$$

can be computed.

More general weighting schemes can be obtained in a straightforward manner. If $y_t = \Delta^k z_t$ and observations on z_t or on $\sum_{i=0}^{m-1} z_{t-i}$, $t \in T_m$, are available, the transformation will always be of type (2).

3. THE IDENTIFICATION OF THE MODEL

To illustrate the nature of the identification problem in dynamic models when observations are missing, we consider a first order autoregressive–second order moving average (ARMA (1, 2)) model, which is a special case of (1) (i.e., when $\beta_k = 0, k = 1, \ldots, K$):

$$y_t = \rho y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \tag{5}$$

with ε_t satisfying the assumptions made for (1). Define $C_l = Ey_t y_{t+l}$.

As the variable y_t is normally distributed, its distribution is determined once the variance and the autocovariances are given. The parameters of model (5) are identified if the system relating the autocovariances C_l, $l = 0, 1, \ldots$ to the parameters in (5) can be solved uniquely for $(\rho, \sigma^2, \theta_1, \theta_2)$, which is the case provided $|\rho| < 1$, the roots of the moving average polynomial lie on or outside the unit circle and are different from $\rho \to -\rho^{-1}$.

If however y_t is observed every second period, only every second autocovariance of y_t can be estimated. The model, which can be written as

$$y_t = \rho^2 y_{t-2} + \varepsilon_t + (\theta_1 + \rho)\varepsilon_{t-1} + (\theta_2 + \rho\theta_1)\varepsilon_{t-2} + \rho\theta_2 \varepsilon_{t-3}, \tag{6}$$

is identified if

$$\begin{aligned} C_0 &= \rho^2 C_2 + \sigma^2[1 + (\theta_1 + \rho)^2 + (\rho^2 + \theta_2 + \rho\theta_1)(\theta_2 + \rho\theta_1) + \rho^2\theta_2^2 \\ &\quad + \rho^3\theta_2(\theta_1 + \rho)], \\ C_2 &= \rho^2 C_0 + \sigma^2[\rho\theta_1 + \theta_2 + \rho^2\theta_2 + \rho\theta_1\theta_2], \\ C_k &= \rho^2 C_{k-2}, \quad k > 2, \end{aligned} \tag{7}$$

can be solved for $(\rho, \sigma^2, \theta_1, \theta_2)$. Eliminating y_{t-1} from (5) to get (6) is in fact equivalent to marginalizing with respect to y_{t-1}. When θ_1 and θ_2 are known to be zero, i.e., y_t is generated by a first order autoregressive model, (7) can simply be solved for ρ^2 and σ^2. Without additional a priori information, the AR (1) model is not identified as no information on the sign of ρ is available.

This finding is at variance with a conclusion by Telser [**14**].[3] The AR (1) model is locally identified, which was implicitly shown (for $\rho \neq 0$) by several authors who established the information matrix of ρ and σ^2 in this model. When the restriction $\rho = 0$ is ignored, although it holds true, the Hessian of the log-likelihood function is singular, but ρ is still identified, even globally. It can be consistently estimated e.g., from the ratio of sample autocovariances $\hat{C}_4$ and $\hat{C}_2$. Convergence of the estimate of ρ however is then slower. For a discussion of the asymptotic distribution of estimators for a locally identified model when the gradient is not of full rank, the reader is referred to Sargan [**13**]. Another interesting conclusion that can be derived from (7) is that pure moving average models of order 1 or

[3] For the AR (1) model with observations for every second period, Telser states that in $\eta_t = y_t - \rho^2 y_{t-2} = \varepsilon_t + \rho\varepsilon_{t-1}$ "all powers of the roots are present" (p. 493). This is not true however for the variance of η_t as Telser implicitly claims below his equation (31).

2 as well as the ARMA (1, 2) model are not identified. The order condition that the MA coefficients have to appear in at least $q+1$ equations of (7) for the model to be identified is not satisfied. These examples are special cases of a more general result that will be presented at the end of this section: when only skipped data are observed (no matter how many periods lie between two successive observations), an ARMA (p, q) model is not identified if $q > p$. That an ARMA (1, 1) model is locally identified if $\rho \neq 0$ can be shown by evaluating the Jacobian of the transformation of the equations for (C_0, C_2, C_4) in (7) to $(\rho, \sigma^2, \theta_1)$, which is $2\rho^2 C_2 \sigma^2 (1-\theta_1^2)/(1-\rho^2)$. The presence of an autoregressive parameter in this example helps to identify the moving average parameter. Notice however that the ARMA (1, 1) model is not globally identified, because if $(\bar{\rho}, \bar{\sigma}^2, \bar{\theta}_1)$ satisfies (7), so does $(-\bar{\rho}, \bar{\sigma}^2, -\bar{\theta}_1)$.

The identification problem clearly shows up in the shape of the log-likelihood function. For an ARMA-model $\rho(L)y_t = \theta(L)\varepsilon_t$, $\sigma^2 = 1$, $t \in T_1$, when observations are skipped denote the data generating process by $\tilde{\rho}(L^m)y_t = \tilde{\theta}(L^m)v_t$, $t \in T_m$, with v_t being a white noise with variance σ_v^2. We evaluate the large sample value of the log-likelihood function

$$f(\hat{\rho}, \hat{\theta}) = -\frac{T}{2m} \ln C(\hat{\rho}, \hat{\theta}) - \frac{T}{2m}$$

at the point $(\hat{\rho}, \hat{\theta})$ with C being the residual variance

$$E(\hat{v}_t^2) = E[\hat{\tilde{\rho}}(L^m)\tilde{\rho}^{-1}(L^m)\hat{\tilde{\theta}}^{-1}(L^m)\tilde{\theta}(L^m)v_t]^2.$$

For instance for an ARMA (1, 1)-process $(1-\rho L)y_t = (1+\theta L)\varepsilon_t$ and $m=2$, we have $\tilde{\rho}(L^2) = 1-\rho^2 L^2$ and $\tilde{\theta}(L^2) = 1+\tilde{\theta}L^2$, with $\tilde{\theta}$ and σ_v^2 obtained from the relations $\tilde{\theta}\sigma_v^2 = \rho\theta$ and $(1+\tilde{\theta}^2)\sigma_v^2 = [1+(\rho+\theta)^2+\rho^2\theta^2]$. (For the ease of exposition, parameter subscripts are deleted.)

We plot $f(\hat{\rho}, \hat{\theta})$ for the values $(\hat{\rho}, \hat{\theta})$ which are not significantly different from the true parameter values (ρ, θ) at which $f(\hat{\rho}, \hat{\theta})$ reaches a maximum and we set $f(\hat{\rho}, \hat{\theta})$ equal to $f(\rho, \theta) - \frac{1}{2}\chi^2_{.95}(n)$, whenever $f(\rho, \theta) - f(\hat{\rho}, \hat{\theta}) \geq \frac{1}{2}\chi^2_{.95}(n)$, with n being the number of parameters in $\rho(L)$ and $\theta(L)$.

From Figure 1, it is quite obvious that the parameters are locally but not globally identified. The value of f does not change if the sign of both $\hat{\rho}$ and $\hat{\theta}$ changes. Moreover f reaches its maximum at $\hat{\rho} = \rho$ and $\hat{\theta} = \theta$ and at $\hat{\rho} = -\rho$ and $\hat{\theta} = -\theta$. Once we have a priori knowledge on the sign of ρ or θ (both being different from zero), the model is globally identified. The two maxima are separated by a set of points, which are significantly different from the true parameter values. The value of θ is not essential for the shape of f. A picture with local maxima similar to Figure 1 also arises when the true model is AR (1), i.e., when $\theta = 0$. Figure 2 is even more instructive. As noticed above, when $\rho = 0$, θ is not identified. The likelihood function is constant for different values of θ, when $\rho = 0$. Notice also that f is constant and reaches a maximum on the set of points $\hat{\rho} = -\hat{\theta}$.

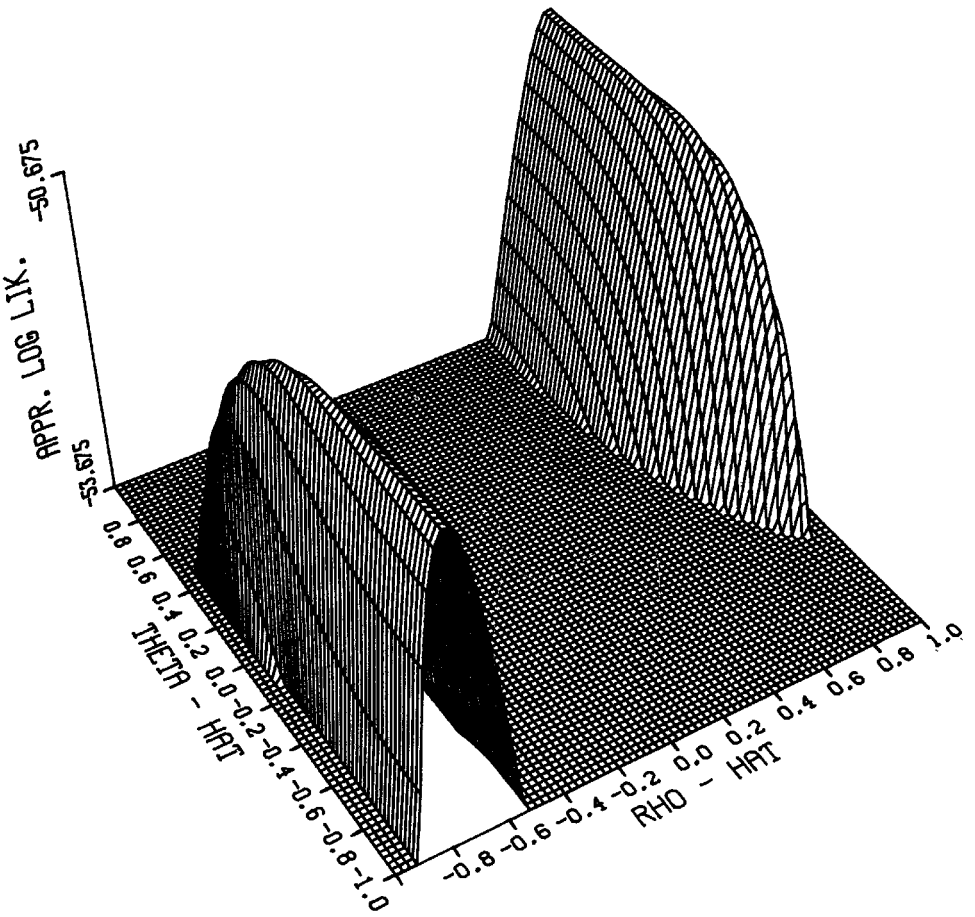

FIGURE 1—ARMA (1, 1)-model: $\rho = .8$, $\theta = .5$, $m = 2$, $T = 100$.

Now we assume that we observe every second period the aggregate over two periods. From (5) we derive the relationship for the temporally aggregated data

$$\bar{y}_t = \rho \bar{y}_{t-1} + \bar{\varepsilon}_t + \theta_1 \bar{\varepsilon}_{t-1} + \theta_2 \bar{\varepsilon}_{t-2}, \tag{5'}$$

where $\bar{y}_t = y_t + y_{t-1}$ and $\bar{\varepsilon}_t = \varepsilon_t + \varepsilon_{t-1}$. For this model, the expression analogous to (6) is

$$\bar{y}_t = \rho^2 \bar{y}_{t-2} + \bar{\varepsilon}_t + (\theta_1 + \rho)\bar{\varepsilon}_{t-1} + (\theta_2 + \rho\theta_1)\bar{\varepsilon}_{t-2} + \rho\theta_2 \bar{\varepsilon}_{t-3}. \tag{6'}$$

Equation (6') can be represented as a first-order autoregressive second-order moving average model in L^2:

$$\bar{y}_t = \rho^2 \bar{y}_{t-2} + (1 - \omega_1 L^2 - \omega_2 L^4) v_t, \tag{7'}$$

where v_t is a white noise with mean zero and variance σ_v^2. It is immediately clear that an MA (2) model for y_t, when $\bar{y}_t$ is observed for $t \in T_2$, is not identified in this case either, as there are only two parameters σ_v^2 and ω_1 in which σ^2, θ_1, and θ_2 appear, when $\rho = 0$ (ω_2 being zero). The MA (1) and ARMA (1, 2) models are identified, provided the usual requirement for the parameters in (7') are satisfied. For instance, when $\rho = \theta_2 = 0$, (7') reduces to an MA (1)-model in L^2, with the

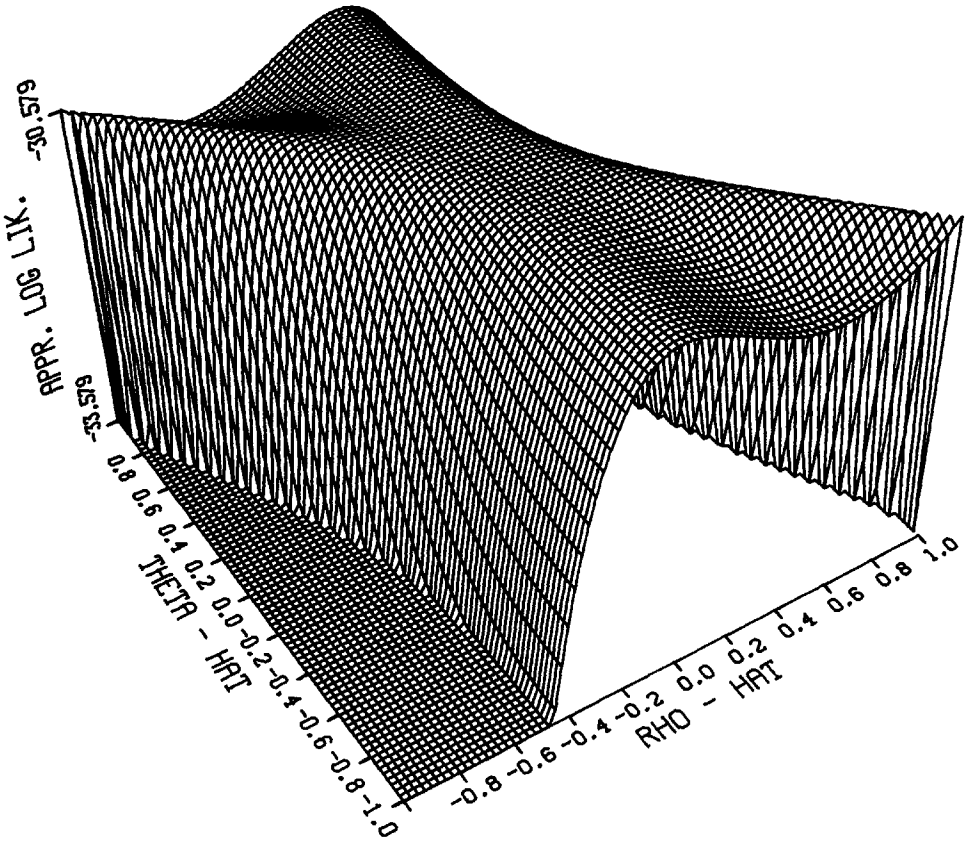

FIGURE 2—MA (1)-model: $\theta = .5$, $m = 2$, $T = 100$.

relations $-\omega_1\sigma_v^2 = \theta_1\sigma^2$ and $(1+\omega_1^2)\sigma_v^2 = [1+(1+\theta_1)^2+\theta_1^2]\sigma^2$ from which σ^2 and θ_1 can be determined. For practical purposes, it can be quite useful to split the identification problem into two parts by first checking the identification of the data generating process (e.g., (7')) and then examining the relation between the parameters of the data generating process and the parameters of interest. Notice finally, that observing sums instead of single realizations helps in identifying the parameters. The sign of ρ is determined here.

To illustrate the impact of the presence of exogenous variables for the parameter identification, we consider the model

$$y_t = \rho y_{t-1} + \beta x_t + \varepsilon_t \tag{8}$$

with ε_t being a normally distributed white noise, and we assume as in (7) that skipped data are available every second period.

Substitution of the model for y_{t-1} yields a model for the observed variables

$$y_t = \rho^2 y_{t-2} + \beta x_t + \beta\rho x_{t-1} + u_t \tag{9}$$

with u_t being a normally distributed white noise with mean zero and variance $\sigma^2(1+\rho^2)$, for $t \in T_2$. It should be noted that if $\beta \neq 0$, the regression coefficients in (9) are odd functions of ρ, so that there is usually information available on

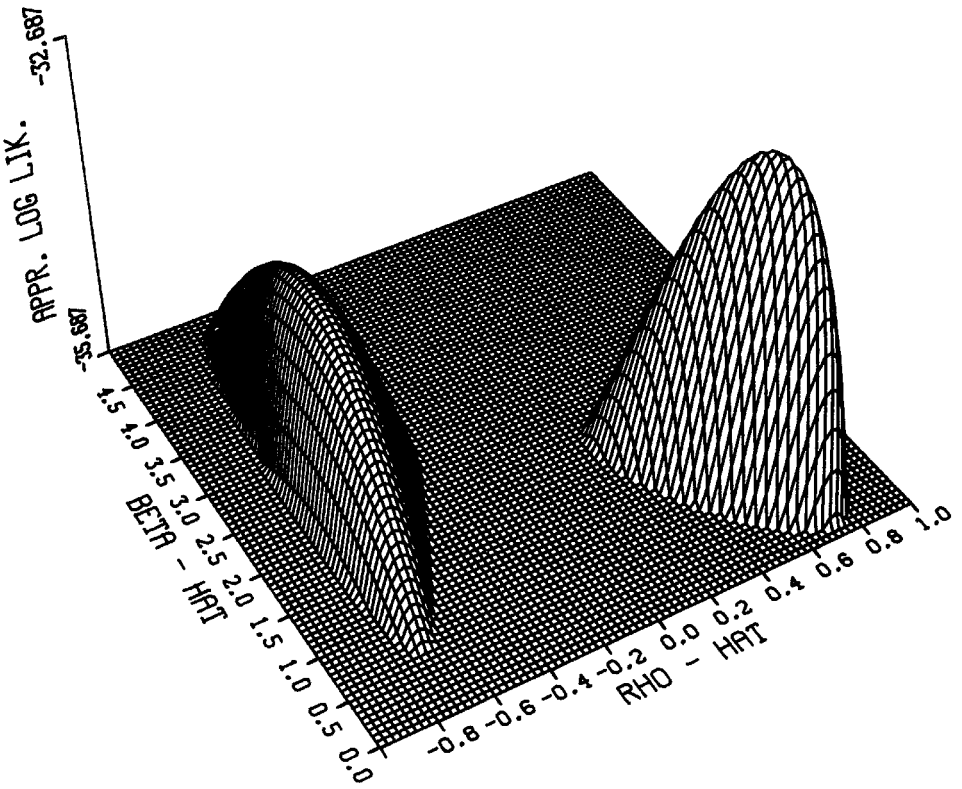

FIGURE 3—Regression model, skipped data: $\rho = .6$, $\beta = 1$, $m = 2$, $T = 100$.

the sign of ρ. The coefficients of (9) are not only identified, they are also subject to restriction, so that the validity of the specification (8) can be tested using (9).

To illustrate the identification problem, we consider the regression model (8). For $\beta = 1$, $\rho = .6$, $\sigma^2 = 1$, $T = 100$ and skipped observations with $m = 2$, the large sample value of the log-likelihood function is plotted in Figure 3. We assume that x_t is generated by a first order autoregressive process with parameter $\gamma = .95$. The value of R^2 is used to determine the variance of x_t. A small value of R^2 ($R^2 = .3$) is chosen to make the figure informative. The log-likelihood function has two local maxima. However, there is only one global maximum which corresponds to the true parameter values. The smallest local maximum will be unimportant for large T, but for small samples, its presence requires some care when the parameters of the model are estimated.

Finally we discuss the identification problem for the general model (1). For this purpose, we use a transformation that has been introduced by Amemiya and Wu [1]. Defining

$$(10) \qquad \tilde{x}_{jt} = w(L)x_{jt} \quad \text{and} \quad \tilde{\varepsilon}_t = w(L)\varepsilon_t,$$

where

$$w(L) = \sum_{i=0}^{A} w_i L^i,$$

we have

$$(11) \qquad \rho(L)\tilde{y}_t = \sum_{k=1}^{K} \beta_k \tilde{x}_{kt} + \theta(L)\tilde{\varepsilon}_t.$$

Now let $\alpha_1, \alpha_2, \ldots, \alpha_p$ be the (possibly complex) roots of the polynomial equation $\rho(L^{-1}) = 0$. Multiplying (11) by

$$\alpha(L) = \prod_{i=1}^{p} (1-\alpha_i L)^{-1}(1-\alpha_i^m L^m) = \prod_{i=1}^{p} \left(\sum_{l=0}^{m-1} \alpha_i^l L^l\right),$$

we get

$$(12) \qquad \prod_{i=1}^{p} (1-\alpha_i^m L^m)\tilde{y}_t = \sum_{k=1}^{K} \alpha(L)\beta_k \tilde{x}_{kt} + \alpha(L)\theta(L)\tilde{\varepsilon}_t.$$

As m is the time lag between subsequent observations on the endogenous variable and because all data on the exogenous variables are assumed to be available, (12) is an expression in observed variables. Introducing a new parametrization for notational convenience and assuming that x_{1t} is the constant term, we can write equation (12) as

$$(13) \qquad \tilde{y}_t = \sum_{i=1}^{p} \Psi_i \tilde{y}_{t-im} + \delta_1 + \sum_{k=2}^{K} \sum_{l=0}^{p(m-1)} \delta_{kl} \tilde{x}_{kt-l} + e_t,$$

where e_t is a MA disturbance term

$$e_t = \sum_{l=0}^{p(m-1)+q+A} \eta_l L^l \varepsilon_t$$

with η_l being defined by

$$\sum_{l=0}^{p(m-1)+q+A} \eta_l L^l = \alpha(L)\theta(L)w(L)$$

and

$$Ee_t e_{t-im} = \xi_i = \sigma^2 \sum_{j=im}^{p(m-1)+q+A} \eta_j \eta_{j-im}.$$

The definition of Ψ_i and δ_{kl} should be clear from (12). Notice also that the parameters in (13) are all real because the α_i's will be in conjugate pairs if they are complex. Equation (13) states that the observations are generated by a dynamic regression model as well. This model is referred to as the transformed difference equation. The order of the MA disturbance in (13) denoted by mq^* satisfies $mq^* \leq (m-1)p + q + A$. It should be noted that when for one exogenous variable only skipped data are available while all other variables are observed at each time period, a transformation similar to that used above can be applied in order to obtain a dynamic regression equation in which only observed variables appear.

The conditional density function for the observed endogenous variables given the exogenous variables can be written in terms of the parameters (Ψ, δ, ξ) which are functions $(\Psi, \delta, \xi) = f(\rho, \beta, \theta)$ of the parameters in (1). Here θ denotes the $g+1$ vector $(\theta_1, \ldots, \theta_g, \sigma^2)$. Alternatively, the disturbance in (13) can be parametrized as a q^*th order MA process $u_t = \sum_{j=0}^{q^*} \omega_j v_{t-jm} = \omega(L^m)v_t$, $\omega_0 = 1$, where v_t is a white noise with mean zero and variance σ_v^2.

A sufficient condition for global identification of (ρ, β, θ) on a subset P of the parameter space is therefore that the parameters (Ψ, δ, ξ) are identified in $f(P)$ without the use of the restrictions on (Ψ, δ, ξ) implied by $f(\rho, \beta, \theta)$ and that the equations $(\Psi, \delta, \xi) = f(\rho, \beta, \theta)$ have a unique solution $(\rho_0, \beta_0, \theta_0)$ in P for every (Ψ, δ, ξ) in $f(P)$. A necessary condition is that $(\Psi, \delta, \xi) = f(\rho, \beta, \theta)$ has a unique solution $(\rho_0, \beta_0, \theta_0)$ for every (Ψ, δ, ξ) in $f(P)$, i.e., that f is injective.

COROLLARY: *The MA parameters θ_j and σ^2 in model* (1) *are not identified if* $q > p + (q - p + A)/m$, *that is if* $q > p + A/(m-1)$.

PROOF: The q θ_j's and σ^2 appear in $f(\rho, \beta, \theta)$ only through the $q^* + 1$ nonzero ξ_i's. The necessary conditions can therefore not be met if $q > p + (q - p + A)/m$.

The corollary implies that the dynamic regression model (1) is not identified when only skipped data are available (transformation pattern (2) and $A = 0$) if $q > p$. Similarly, it is not identified when only aggregates are available (transformation pattern (2) and $A = m - 1$) if $q > p + 1$. Of course some parameters in model (1) can be identified, although other parameters are not. This is a problem of aliasing. Writing

$$1 + \sum_{i=1}^{p} \Psi_i L^{im} = \prod_{i=1}^{p} (1 - \tilde{\Psi}_i L^m)$$

we can see that the ρ_i's are identified if there is sufficient a priori information to determine them uniquely from $\alpha_i^m = \tilde{\Psi}_i$.

If $K > 0$, this a priori information is not always needed as the roots of $\alpha_i^m = \tilde{\Psi}_i$ can partly be determined from the δ_{kl}. If the vector β is identified and there is at least one $\tilde{x}_{kt}$ such that the variables $\tilde{x}_{kt-j}$, $j = 0, \ldots, p(m-1)$ are linearly independent, $\alpha(L)$ can be obtained from the δ_{kl} and $\rho(L)$ can be obtained by dividing $\prod_{i=1}^{p} (1 - \alpha_i^m L^m)$ by $\alpha(L)$.

Finally, it may be difficult to accurately estimate the $\tilde{\Psi}_i$'s and $\alpha(L)$ when m or p are large as will be seen in the next section. The identification of β is straightforward if (13) is identified. A necessary condition for the identification of β is that the K variables $\alpha(L)\tilde{x}_{tk}$ are linearly independent. For instance, β is not identified in models with a constant term and a seasonal dummy over m periods or with two dummies because $\alpha(L)\tilde{x}_{kt} = \lambda\alpha(L)\tilde{x}_{k't}$ with $k \neq k'$ in that case; that is β_k and $\beta_{k'}$ enter in (12) only as $(\beta_k + \lambda\beta_{k'})$, hence they cannot be identified. Finally when $q \leq p + A/(m-1)$, the identification of θ can be checked by showing that the Jacobian of the transformation $\xi = \xi(\theta)$ has full rank and that θ is a regular point (see, e.g., Rothenberg [**12**]).

4. INFORMATION LOSS DUE TO INCOMPLETE SAMPLING

In the preceding section, we have seen that when data are incomplete large areas of the parameter space may not differ significantly from the true parameter values. Moreover, local maxima will often occur and the Hessian matrix may become singular for some points in the parameter space. The examples given in

Section 3 indicate that identification and the related problem of estimation require additional care when the sample is incomplete. We shall see that in some cases, almost unidentifiability occurs, although formally the model is identified. In this section, we examine the loss of information due to missing data by comparing the asymptotic efficiency of the ML estimator of the parameters in a dynamic regression model, when the complete sample is available, with that of the ML estimator, when some observations on the endogenous variable are missing.[4] The relative efficiency is perhaps of limited value for practical situations, as an investigator usually does not have the choice between either using a complete sample or relying on incomplete data. However, the results that we shall present are interesting for the following reasons.

For empirical work, it is important to know that for some parameters the large sample precision of the ML estimator deteriorates dramatically as a result of incomplete data, whereas other parameters can be estimated fairly accurately in large but incomplete samples. Using the information matrix, we show which restrictions become essential for identification and accurate estimation when observations are missing. Other consistent, but not fully efficient estimators will have a still larger asymptotic variance. The asymptotic efficiency of alternative consistent estimators has been investigated by Palm and Nijman [**10, 11**]. Finally the results indicate for which kind of sampling and for which parameter values the loss of efficiency due to missing observations is important. This may be of interest to those in charge of data collection so that they can better appreciate which gain can be expected from a more detailed data collection.

In Tables I and II, we give the ratio of the asymptotic variance of the ML estimator of the parameters for skipped data, aggregate observations and for $w(L)=(1+L+\cdots+L^{m-1})^2$, with respect to that when complete data are available. The transformation $w(L)=(1+L+\cdots+L^{m-1})^2$ occurs when the change of a variable z_t, $y_t=z_t-z_{t-1}$, is explained in the regression model (1), whereas one observes an aggregate of z_t, or when the model (1) explains second differences of skipped observations. Table I contains results for univariate ARMA (1, 1)-models. In Table II, we report on the relative efficiency for regression models. The number of periods m is 2, 3, and 4 respectively and is given in the first column of the tables. The true parameter values are $\rho\in\{-.8, 0, .8\}$, $\sigma^2=1$, $\theta\in\{-.6, 0, .6\}$. The coefficient β equals one in Table II, where several alternative processes are considered for the exogenous variable:

$x_t=u_t$ is a white noise with mean zero and variance σ_u^2, denoted by WHI;

$x_t=.9x_{t-1}+u_t$, denoted by AR;

$x_t=\mu t$, a trend denoted by TRE;

$x_t=x_{t-1}+u_t$, a random walk denoted by RWA;

$x_t=x_{t-1}+u_t+\mu$, a random walk with drift denoted by RWD, with $\mu/\sigma_u=1$.

[4] An alternative measure, proposed by Dempster et al. [**3**], is the information matrix associated with the density for the missing observations conditionally on the data. Notice that the expectation of their measure with respect to the data equals the difference between the information matrices of the complete sample and the incomplete sample respectively.

TABLE I

ASYMPTOTIC EFFICIENCY COMPARISONS FOR UNIVARIATE ARMA-MODELS

				Skipped Observ.			Aggregates			$w(L)=(1+L+\cdots+L^{m-1})^2$		
m	Model	ρ	θ	ρ_{ML}	θ_{ML}	σ^2_{ML}	ρ_{ML}	θ_{ML}	σ^2_{ML}	ρ_{ML}	θ_{ML}	σ^2_{ML}
2	AR	−0.8	0.0	1.3		2.2	36.0		3.7	5.8*3		1.1*3
2	AR	0.0	0.0	INF		3.0	8.0		6.0	8.2		7.7
2	AR	0.8	0.0	1.3		2.2	1.3		2.3	1.3		2.3
3	AR	−0.8	0.0	1.7		4.0	1.9		3.4	2.3		3.1
3	AR	0.0	0.0	INF		INF	27.0		27.0	42.4		55.6
3	AR	0.8	0.0	1.7		4.0	1.6		4.1	1.7		4.1
4	AR	−0.8	0.0	2.2		6.3	80.8		4.6	6.7		5.9
4	AR	0.0	0.0	INF		INF	64.0		76.0	1.5*2		2.4*2
4	AR	0.8	0.0	2.2		6.3	2.0		6.9	2.3		6.7
2	MA	0.0	−0.6		NID	NID		1.9	2.2		1.6	2.6
2	MA	0.0	0.0		NID	NID		8.0	6.0		8.2	7.7
2	MA	0.0	0.6		NID	NID		3.8*2	1.5*2		4.1*2	1.7*2
3	MA	0.0	−0.6		NID	NID		3.0	4.1		2.7	5.9
3	MA	0.0	0.0		NID	NID		27.0	27.0		42.4	55.6
3	MA	0.0	0.6		NID	NID		2.1*3	9.2*2		4.0*3	1.8*3
4	MA	0.0	−0.6		NID	NID		4.2	6.6		4.4	12.6
4	MA	0.0	0.0		NID	NID		64.0	76.0		1.5*2	2.4*2
4	MA	0.0	0.6		NID	NID		6.2*3	2.8*3		1.8*4	8.7*3
2	ARMA	−0.8	−0.6	2.2	6.8*2	2.8*2	3.1	7.5	3.2			
2	ARMA	−0.8	0.0	2.0	13.1	8.8	28.9	13.7	6.0			
2	ARMA	−0.8	0.6	2.8	4.0	2.2	1.3*3	1.1*2	1.3*3			
2	ARMA	0.0	−0.6	INF	NID	NID	INF	INF	INF			
2	ARMA	0.0	0.6	INF	NID	NID	INF	INF	INF			
2	ARMA	0.8	−0.6	2.8	4.0	2.2	1.4	1.8	2.9			
2	ARMA	0.8	0.0	2.0	13.1	8.8	1.5	10.1	10.1			
2	ARMA	0.8	0.6	2.2	6.8*2	2.8*2	1.8	5.9*2	2.6*2			

NOTE: By "$y*z$," we denote the figure $y\times 10^z$.

The values of the coefficient of determination R^2 which are close to those frequently observed in empirical work with economic time series, have been used to determine the parameter of the process for x_t. The large sample variances needed for the efficiency comparisons in the tables have been computed as the diagonal elements of the inverse of the information matrix (for some details, see the Appendix). When the process for x_t is nonstationary, its parameters are determined from R^2 with $T=30$.

The reader should have a look at the tables. By NID, we indicate that the parameter is not locally identified. INF denotes that the asymptotic variance of the ML estimator is infinite due to a singularity of the Hessian matrix. When results are reported for a parameter that is zero, this means that the parameter is considered as being unknown. If no result is reported for an individual parameter, it is assumed to be zero. For ρ, σ^2, and θ, a large relative efficiency of the ML estimator for complete data is caused by the large variance of the ML estimator for incomplete data. For β, the relative efficiency of the former is sometimes important, although the variance for the estimator based on incomplete data seems to be reasonable. We checked this by computing numerator and denominator of the relative efficiency separately. In order to give some insight in the relationship between the results in the tables and the characteristics of the model, we derived analytical expressions for the relative efficiency for some simple models. For instance, for the pure first order autoregressive model, the relative efficiency of the ML estimator for ρ compared with that for skipped data ($m=2$) is equal to $(1+\rho^2)/2\rho^2 \geq 1$ and that for σ^2 is $(3+\rho^2)/(1+\rho^2)$. Then we get the following results:

ρ	0	.1	.3	.5	.7	.9
rel. eff. ρ_{ML}	INF	50.5	6.05	2.50	1.52	1.12
rel. eff. σ^2_{ML}	3.00	2.98	2.83	2.60	2.34	2.10

From these figures, it is obvious that the relative efficiency of ρ_{ML} is very sensitive to the value of ρ, whereas that for σ^2_{ML} increases only slowly with decreasing ρ.

The efficiency of the ML-estimate of ρ is infinite when $\rho=0$. For σ^2, the efficiency of the ML estimate is finite. Both parameters ρ and σ^2 are globally identified when $\rho=0$ and can be estimated consistently. Notice that $(0, \sigma^2)$ is not a regular point of the Hessian matrix, i.e., a point in the parameter space for which there exists an open neighborhood in which the Hessian matrix has constant rank. Therefore, Rothenberg's [**12**] necessary and sufficient condition for local identification requiring the Hessian to be nonsingular does not apply. For a regression model with x_t generated by a linear trend, $\rho=0$ and with skipped observations ($m=2$), the relative efficiency for ρ_{ML} is $2+8\sigma^2/\mu^2$. Using the value of R^2 with $T=30$ and 100 respectively to determine σ^2/μ^2, we obtain the following results:

R^2	.7		.95		.99	
T	30	100	30	100	30	100
rel. eff. ρ_{ML}	1082.56	1450.07	134.65	178.08	27.47	34.17

TABLE II

ASYMPTOTIC EFFICIENCY COMPARISONS FOR REGRESSION MODELS

						Skipped Observ.				Aggregates				$w(L)=(1+L+\cdots+L^{m-1})^2$		
m	Model	x_t	R^2	ρ	θ	ρ_{ML}	θ_{ML}	β_{ML}	σ^2_{ML}	ρ_{ML}	θ_{ML}	β_{ML}	σ^2_{ML}	ρ_{ML}	β_{ML}	σ^2_{ML}
2	ARX	WHI	0.95	−0.8	0.0	1.2		2.1	2.0	3.7		2.1	2.0	2.2	1.2	2.0
2	ARX	WHI	0.95	0.0	0.0	2.1		2.0	2.0	2.8		2.7	2.1	1.8	1.7	2.1
2	ARX	WHI	0.95	0.8	0.0	1.2		2.1	2.0	1.2		2.1	2.0	1.0	1.5	2.0
2	ARX	AR	0.95	−0.8	0.0	1.2		7.0	2.0	3.8		2.9	2.0	2.2	1.8	2.0
2	ARX	AR	0.95	0.0	0.0	2.6		2.4	2.0	2.4		2.3	2.3	1.7	1.7	2.3
2	ARX	AR	0.95	0.8	0.0	1.1		1.2	2.0	1.1		1.1	2.0	1.0	1.0	2.0
2	ARX	RWA	0.95	−0.8	0.0	1.2		1.2	2.1	4.4		4.4	2.1	2.7	2.7	2.1
2	ARX	RWA	0.95	0.0	0.0	3.6		3.6	2.0	3.0		3.0	2.7	2.3	2.3	2.7
2	ARX	RWA	0.95	0.8	0.0	1.1		1.1	2.0	1.1		1.1	2.0	1.0	1.0	2.0
2	ARX	RWD	0.95	−0.8	0.0	1.2		1.2	2.2	10.8		10.8	2.4	6.3	6.3	2.8
2	ARX	RWD	0.95	0.0	0.0	29.8		29.8	2.0	6.4		6.4	5.0	6.1	6.1	5.9
2	ARX	RWD	0.95	0.8	0.0	1.1		1.1	2.1	1.1		1.1	2.1	1.1	1.1	2.1
2	ARX	TRE	0.95	−0.8	0.0	1.3		1.3	2.2	44.9		44.9	4.1	1.3*3	1.3*3	2.4*2
2	ARX	TRE	0.95	0.0	0.0	1.3*2		1.3*2	2.0	7.2		7.2	5.6	7.4	7.4	7.0
2	ARX	TRE	0.95	0.8	0.0	1.2		1.2	2.1	1.1		1.1	2.1	1.1	1.1	2.1
3	ARX	AR	0.95	−0.8	0.0	1.3		3.9	3.1	1.6		1.5	3.0	1.1	1.2	3.0
3	ARX	AR	0.95	0.0	0.0	3.8		3.6	3.0	4.0		3.8	3.8	5.5	5.2	4.5
3	ARX	AR	0.95	0.8	0.0	1.3		1.4	3.1	1.2		1.3	3.1	2.0	2.1	3.2
3	ARX	RWD	0.95	−0.8	0.0	1.5		1.5	3.6	1.8		1.8	3.3	1.7	1.7	3.0
3	ARX	RWD	0.95	0.0	0.0	44.8		44.8	3.0	15.2		15.2	15.6	24.8	24.8	31.6
3	ARX	RWD	0.95	0.8	0.0	1.2		1.2	3.2	1.2		1.2	3.3	1.8	1.8	3.4
4	ARX	AR	0.95	−0.8	0.0	1.6		12.6	4.2	5.3		3.9	4.0	7.6	6.7	4.3

4	ARX	AR	0.95	0.0	0.0	5.1		4.9	4.0	5.6		5.3	5.4	12.9	12.3	8.3
4	ARX	AR	0.95	0.8	0.0	1.5		1.6	4.2	1.3		1.4	4.2	3.3	3.6	4.5
4	ARX	RWD	0.95	−0.8	0.0	1.9		1.9	5.4	13.8		13.8	4.1	7.0	7.0	5.4
4	ARX	RWD	0.95	0.0	0.0	59.7		59.7	4.0	24.2		24.2	29.4	67.0	67.0	99.4
4	ARX	RWD	0.95	0.8	0.0	1.4		1.4	4.5	1.3		1.3	4.6	2.8	2.8	5.1
2	ARMAX	AR	0.95	−0.8	−0.6	1.5	4.1*2	10.1	1.6*2	1.8	7.5	1.8	2.8			
2	ARMAX	AR	0.95	−0.8	0.0	1.3	8.6	6.8	6.1	3.6	16.4	2.8	2.7			
2	ARMAX	AR	0.95	−0.8	0.6	1.4	2.7	3.2	2.2	22.8	6.0*2	5.1	1.9*2			
2	ARMAX	AR	0.95	0.0	−0.6	3.8	NID	3.9	NID	1.7	1.8	1.6	2.2			
2	ARMAX	AR	0.95	0.0	0.0	2.0	NID	2.0	NID	2.1	6.7	2.1	6.0			
2	ARMAX	AR	0.95	0.0	0.6	3.8	NID	3.8	NID	5.6	3.8*2	5.4	1.6*2			
2	ARMAX	AR	0.95	0.8	−0.6	1.4	2.6	1.4	2.2	1.0	1.6	1.0	2.6			
2	ARMAX	AR	0.95	0.8	0.0	1.2	8.6	1.2	6.1	1.1	8.3	1.1	7.8			
2	ARMAX	AR	0.95	0.8	0.6	1.4	4.1*2	1.6	1.6*2	1.4	4.2*2	1.6	1.7*2			
2	ARX	AR	0.70	−0.8	0.0	1.2		9.5	2.1	6.4		3.6	2.2	4.1	2.4	2.4
2	ARX	AR	0.70	0.0	0.0	6.5		4.6	2.0	4.2		3.1	3.5	3.3	2.5	3.6
2	ARX	AR	0.70	0.8	0.0	1.2		1.2	2.1	1.2		1.1	2.1	1.1	1.1	2.1
2	ARX	RWD	0.70	−0.8	0.0	1.3		1.3	2.2	24.0		24.0	3.1	39.0	39.0	8.8
2	ARX	RWD	0.70	0.0	0.0	2.3*2		2.3*2	2.0	7.7		7.7	5.8	7.8	7.8	7.4
2	ARX	RWD	0.70	0.8	0.0	1.2		1.2	2.2	1.2		1.2	2.2	1.2	1.2	2.2
2	ARMAX	AR	0.70	−0.8	−0.6	1.7	4.9*2	15.9	1.9*2	2.2	7.4	2.2	2.9			
2	ARMAX	AR	0.70	−0.8	0.0	1.4	9.7	8.6	6.8	5.3	14.2	3.4	2.9			
2	ARMAX	AR	0.70	−0.8	0.6	1.5	2.5	3.2	2.2	24.6	4.3*2	5.1	1.9*2			
2	ARMAX	AR	0.70	0.0	−0.6	5.5	NID	5.3	NID	2.2	1.5	2.2	2.3			
2	ARMAX	AR	0.70	0.0	0.0	2.0	NID	2.0	NID	2.1	3.9	2.1	6.0			
2	ARMAX	AR	0.70	0.0	0.6	5.5	NID	4.4	NID	7.6	4.4*2	6.1	2.2*2			
2	ARMAX	AR	0.70	0.8	−0.6	1.4	2.5	1.4	2.2	1.0	1.6	1.1	2.6			
2	ARMAX	AR	0.70	0.8	0.0	1.4	9.6	1.3	6.8	1.2	8.7	1.2	8.4			
2	ARMAX	AR	0.70	0.8	0.6	1.6	4.9*2	1.7	1.9*2	1.5	4.8*2	1.6	2.0*2			

Notice that for trending x_t, the loss of efficiency caused by skipped data remains large, even for values of R^2 close to 1. From the results in the tables, it becomes obvious that for skipped data, a large loss of efficiency occurs when the model is overfitted, e.g., if ρ is estimated when it is zero, and when the true value of ρ is close to that of θ. Notice also that for skipped data, the efficiency of ρ and θ is not affected, when the sign of both is changed. The small differences in the results are due to rounding errors. For aggregates, a large efficiency loss arises when θ is large, independently of the value of ρ. For $w(L) = (1 + L + \cdots + L^{m-1})^2$, a large imprecision is found when ρ is negative. In general, the loss of information increases as R^2 decreases. Notice finally that a large inefficiency can be seen as almost unidentifiability of the corresponding parameter.

To conclude, the results in the tables give an indication about the loss of precision in parameter estimates and about the order of magnitude of the variance of the ML estimator when observations are missing. It should be obvious from these results that the loss of precision is not simply proportional to the number of missing data points but that it can strongly depend on the true values of the parameters and on the sampling scheme. In principle, it is possible to express the relative efficiency as a function of the characteristics of the data generating process and the sampling scheme. Except for simple models such as discussed above, the derivation of the formulae for the relative efficiency is tedious and the formulae are complicated so that the insight they give is limited.

5. PARAMETER ESTIMATION FROM INCOMPLETE DATA

As noted in the introduction, parameter estimation has received much attention in the literature. We shall give new results on the properties of a procedure frequently used in applied work and we shall briefly discuss some alternative estimation methods.

Quite often in empirical work, values for the missing observations are obtained by interpolation such that the resulting series is plausible according to some criterion and is in agreement with the observed values of the series. Then the constructed series is used as realization for the missing observations. This procedure has the advantage of being straightforward to apply. However, it yields inconsistent parameter estimates in most occasions. In order to know whether this drawback is relevant for applied work, we shall now investigate the large sample bias of the OLS estimator for ρ and β in (1) using interpolated data.

Boot, Feibes, and Lisman [2] have proposed an interpolation method which has been applied on a large scale. Other more sophisticated methods have also been proposed in the literature. Generalizing their method to other cases than observed aggregates, one obtains the interpolated series $\hat{y}_t$ as the solution to the following optimization problem:

$$\min_{\hat{y}_t} \sum_{t=1}^{T} [(1-L)^d \hat{y}_t]^2 \tag{14}$$

subject to $y = \tilde{y}_t$, for $t \in T_m$, and d being a priori given. Boot et al. [2] suggest using $d = 1$ or $d = 2$. The procedure reflects the fact that many economic time

series are smooth and that the constructed series should have that property too. The interpolated series can be written as a linear transformation $\hat{y} = Ry$ of the realizations, where $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_T)'$, $y = (y_1, \ldots, y_T)'$, and R is a matrix of constants that does not depend on the observations $\tilde{y}_t$. When $q = 0$, ordinary least squares estimates of ρ and β in (1) using the constructed series $\hat{y}_t$ are given by

$$\begin{pmatrix} \dot{\rho} \\ \dot{\beta} \end{pmatrix} = (Z'Z)^{-1}Z'R_0y, \tag{15}$$

where $Z = (R_1y, R_2y, \ldots, R_py, X)$ with R_i being the matrix obtained by deleting the first $(p-i)$ and the last i rows of R. When R is a block-Toeplitz matrix with a limited number of nonzero elements, i.e., with i,jth block equal to A_{j-i} for $|j-i| \leq n$, for some finite n and zero otherwise, we can compute the probability limit of the OLS estimator in (15). Thereby, we use the property that products of matrices of this structure are again block-Toeplitz matrices. The matrix R for the procedure of Boot et al. [2] is not of the form described above but can be very closely approximated by such a matrix. The probability limit of the OLS estimator in (15) can be obtained straightforwardly by replacing the cross-products by their second order moments expressed in terms of the parameters using results of the Appendix.

In Table III, we report the probability limit of $\dot{\rho}$ and $\dot{\beta}$ when aggregates over m periods are observed and the method proposed by Boot et al. [2] is used with $d = 2$ and $m = 4$. (Results for other interpolation schemes can be obtained on request from the authors.) For the exogenous variable x_t, we consider the processes that have been used in Section 4. For the parameters in (1), we choose the values $\rho \in \{-.8, -.4, 0, .4, .8\}$, $\beta = 1$, $\sigma^2 = 1$. For the details, we refer to the preceding section.

From Table III, it is obvious that the probability limit can substantially differ from the true values of ρ and β. When $\rho = .8$, the probability limit of $\dot{\rho}$ is reasonably close to the true value, except for some nonstationary models. For

TABLE III

THE PROBABILITY LIMIT OF THE OLS ESTIMATOR WHEN USING INTERPOLATED DATA FOR THE MISSING AGGREGATE OBSERVATIONS, $m = 4$, $\beta = 1$, $\sigma^2 = 1$

		$\rho = -.8$		$\rho = -.4$		$\rho = 0$		$\rho = .4$		$\rho = .8$	
R^2	x_t	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$	$\dot{\rho}$	$\dot{\beta}$
0.70	WHI	0.86	0.02	0.86	0.03	0.86	0.06	0.87	0.11	0.93	0.20
0.95	WHI	0.86	0.02	0.85	0.03	0.85	0.06	0.86	0.12	0.93	0.20
0.70	AR	0.78	0.11	0.77	0.16	0.78	0.22	0.81	0.34	0.90	0.60
0.95	AR	0.75	0.13	0.71	0.19	0.68	0.30	0.71	0.48	0.85	0.75
0.70	TRE	−0.01	0.56	0.00	0.71	0.02	0.98	0.03	1.61	0.05	4.77
0.95	TRE	−0.01	0.56	−0.01	0.72	−0.01	1.01	−0.00	1.68	−0.00	5.00
0.70	RWA	0.65	0.19	0.72	0.20	0.76	0.24	0.81	0.32	0.87	0.65
0.95	RWA	0.55	0.25	0.54	0.32	0.56	0.44	0.63	0.62	0.79	1.03
0.70	RWD	−0.01	0.56	0.00	0.71	0.02	0.98	0.03	1.62	0.02	4.92
0.95	RWD	−0.01	0.56	−0.01	0.72	−0.01	1.01	−0.01	1.68	−0.00	5.01

most models with nonpositive ρ, OLS largely overestimates ρ. The coefficient β is usually underestimated, except when x_t is nonstationary and $\rho > 0$. The figures slightly improve when R^2 increases. When x_t is generated by a trend or a random walk with drift, the probability limit of $\hat{\rho}$ hardly varies with the true value of ρ. Also, we like to note that the convergence of OLS to its limiting value appeared to be very slow in simulations of the path of OLS estimates as a function of sample size.

Finally, we point out that Boot et al. [2] originally designed their method for economic variables, for which the roots of the autoregressive part are often positive. Therefore, we do not want to draw strong conclusions from the results in Table III, when $\rho = -.8$, $\rho = -.4$, and x_t is a white noise. Nevertheless, as an overall conclusion, we cannot recommend applying OLS to interpolated data, if the aim is to estimate the parameters of a dynamic regression model from an incomplete sample. In the light of this result and the conclusions on the identification problem, we advocate the use of methods which rely on relevant a priori information to analyze regression models with missing observations.

Fortunately, there are many ways for implementing ML and other nearly efficient estimation methods. After the elimination of the unobserved variables, one obtains a dynamic regression model of the form (13) with moving average errors or, when $\beta_k = 0$, $k = 2, \ldots, K$, a univariate ARMA-model. The parameters of these models are usually subject to restrictions implied by $(\Psi, \delta, \xi) = f(\rho, \beta, \theta)$. One way to obtain ML estimates of ρ, β and θ consists in estimating the unrestricted model (13) by means of nonlinear least squares or ML and then apply the method of asymptotic least squares proposed by Gouriéroux et al. [6]. Alternatively, the restricted model (12) can be estimated directly by means of e.g. the Gauss–Newton algorithm using e.g., the chain rule to compute the partial derivatives of the disturbance v_t with respect to the parameters (see Appendix). Some ML procedures do not require explicit marginalization with respect to the unobserved variables. For instance, the log-likelihood function in prediction error decomposition form and its derivatives can be evaluated by means of the Kalman filter in order to obtain ML estimates of the parameters and predictions for the missing data points (see, e.g., Harvey and McKenzie [8]). Alternatively, the EM algorithm proposed by Dempster et al. [3] can be applied to the joint process for data and missing observations to get the ML estimates of the parameters and predictions for the missing observations. As the EM algorithm is based on a sufficient statistic, it is less suited for models with MA parameters. Advantages of the EM algorithm are that the value of the likelihood function increases at each step of iteration and that it moves quickly to a region close to the maximum. Watson and Engle [15] show how the EM algorithm can be implemented by means of the Kalman filter and how standard errors for the ML estimates can be computed. The prediction error decomposition and the EM procedure are also suited for other sampling schemes not discussed here.

Finally, many consistent but not fully efficient estimation methods are available. For instance, besides efficient estimators, Dunsmuir and Robinson [5] present consistent frequency domain moment estimators for ARMA models with ran-

domly missing observations. Palm and Nijman [11] compare the efficiency of several consistent generalized moment estimators which are computationally attractive. These estimators can be fairly efficient compared with the ML estimator provided the most important restrictions are taken into account.

6. CONCLUDING REMARKS

In this paper, we have considered the problems of identification and estimation arising in dynamic regression models and univariate ARMA models, when some realizations of the endogenous variable are not observed. After the presentation of different schemes in which information on the endogenous variable may be available, we consider the conditions for identification of the parameters in regression model with incomplete data.

The examples discussed above show that the data may not be very informative on the parameter values. The Hessian of the log-likelihood function can become singular when some restrictions on the parameters are ignored although the model is still identified. The results on the loss of information due to an incomplete data set indicate that some parameters can be determined with reasonable accuracy whereas other parameters are almost unidentified. Ad-hoc estimation procedures based on interpolated data which do not take account of the features of the model can be heavily biased in large samples. However fortunately, many procedures to obtain ML estimates are available. Each of them has specific computational advantages.

In the light of our results, we should like to advise an investigator to carefully check the identification of the parameters of the dynamic regression model with incompletely observed endogenous variable and to estimate the parameters by ML. Moreover, as some coefficients in the model cannot be determined very accurately from the sample information, we also like to advise him to use all available reliable a priori information.

Vrije Universiteit, Amsterdam, The Netherlands

Manuscript received October, 1982; final revision received March, 1984.

APPENDIX

THE LARGE SAMPLE COVARIANCE MATRIX OF ML ESTIMATES

In this Appendix, we shall briefly outline how the asymptotic covariance matrices for the ML estimator of the parameters in equation (12) have been obtained. The large sample covariances have been used to compute the results in Tables I to II in Section 4. First, we obtain the Hessian matrix of the unrestricted regression model (13). For the ease of the exposition, we assume that there is only one explanatory variable $\tilde{x}_t$ in equation (13) which we write as

$$\Psi(L^m)\tilde{y}_t = \delta(L)\tilde{x}_t + \omega(L^m)v_t,\ v_t \sim N(0, \sigma_v^2). \tag{A.1}$$

The Hessian matrix of the log-likelihood function L with respect to the vector of unrestricted parameters $\nu=(\Psi', \delta', \omega')'$ in (A.1) is

$$\text{(A.2)} \qquad \frac{\partial^2 L}{\partial \nu_i \partial \nu_j} = -\sigma_v^{-2} \sum_{t \in T_m} \left(\frac{\partial v_t}{\partial \nu_i} \frac{\partial v_t}{\partial \nu_j} + \frac{\partial^2 v_t}{\partial \nu_i \partial \nu_j} v_t \right).$$

For the restricted model (12), the information matrix is then given as

$$\text{(A.3)} \qquad -EB \frac{\partial^2 L}{\partial v \, \partial v'} B' = \frac{T}{m} BCB',$$

where $B' = \partial \mu / \partial \varphi'$ with $\mu' = (\nu', \sigma_v^2)$ and $\varphi' = (\rho', \beta', \theta', \sigma^2)$ and

$$C = \begin{bmatrix} C_1 & 0 \\ 0 & 2\sigma_v^4 \end{bmatrix} \quad \text{with} \quad C_1 = \sigma_v^{-2} E \frac{\partial v_t}{\partial \nu} \frac{\partial v_t}{\partial \nu'}.$$

The elements of C_1 can be obtained in a fairly straightforward way (when $\omega(L)$ is invertible) by using results given below together with the following relationships which express the gradients as linear functions of disturbances and explanatory variables:

$$\text{(A.4)} \qquad \frac{\partial v_t}{\partial \Psi_i} = -\omega^{-1}(L^m) \tilde{y}_{t-im} = -\Psi^{-1}(L^m)[v_{t-im} + \omega^{-1}(L^m)\delta(L)\tilde{x}_{t-im}],$$

$$\frac{\partial v_t}{\partial \omega_i} = -\omega^{-1}(L^m) v_{t-im}, \qquad \frac{\partial v_t}{\partial \delta_i} = -\omega^{-1}(L^m)\tilde{x}_{t-i}.$$

The Jacobian matrix B can be obtained by analytical or numerical differentiation.

In order to compute the elements of C_1, we need expressions for the sample moments for the variables y_t and x_t. We assume that x_t is nonstationary (when x_t is stationary, the derivation is straightforward) and generated by $x_t = x_{t-1} + \mu + u_t$, with u_t being independently distributed as $N(0, \sigma_u^2)$ and independent of past x_t's. The process for x_t and y_t can be written as:

$$\text{(A.5)} \qquad x_t = \mu t + \sum_{s=1}^{t} u_s + x_0,$$

$$\text{(A.6)} \qquad y_t = \sum_{i=0}^{t} \rho^i (x_{t-i} + \varepsilon_{t-i}),$$

where x_0, y_{-1}, and ε_0 are assumed to be zero. As $\sum_{t=0}^{T/n} \rho^i i^a$ is $O(T^0)$ for finite a, we ignore the terms of this form. When divided by T, they become negligible in large samples. We denote by "$\doteq$" that the equality holds except for terms of order $O(T^0)$. From (A.5), the second moment for x_t is:

$$\text{(A.7)} \qquad E \sum_{t=0}^{T/n} x_{tn+k} x_{tn+l} \doteq \mu^2 D + \frac{\sigma_u^2}{2n} T^2 + \left(\frac{1}{2} + \frac{r}{n} \right) \sigma_u^2 T,$$

where

$$\text{(A.8)} \qquad D \doteq \sum_{t=0}^{T/n} (nt+k)(nt+l) = \frac{1}{3n} T^3 + \frac{1}{2} \left(1 + \frac{l+k}{n} \right) T^2 + \left[\frac{n}{6} + \frac{1}{2}(l+k) + \frac{lk}{n} \right] T$$

and $r = \min(k, l)$. Using (A.6) and (A.8), the sample moments $E \sum_{t=0}^{T/n} y_{tn+k} y_{tn+l}$ and $E \sum_{t=0}^{T/n} y_{tn+k} x_{tn+l}$ can be expressed as functions of T, n, k, l and the parameters in (A.5) and (A.6).

Finally, the coefficient of determination R^2 defined as the ratio of $E \sum_{t=0}^{T} x_t^{*2} / E \sum_{t=0}^{T} y_t^2$, where $x_t^* = \sum_{i=0}^{t} \rho^i x_{t-i}$, is used to determine the parameter values for the models in Sections 4 and 5.

REFERENCES

[1] AMEMIYA, T., AND R. Y. WU: "The Effect of Aggregation on Prediction in the Autoregressive Model," *Journal of the American Statistical Association*, 67 (1972), 628–632.

[2] BOOT, J. C. G., W. FEIBES, AND J. H. C. LISMAN: "Further Methods of Derivation of Quarterly Figures From Annual Data," *Applied Statistics*, 16 (1967), 65–75.

[3] DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN: "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B, 39 (1977), 1–38.
[4] DUNSMUIR, W.: "Estimation for Stationary Time Series when Data are Irregularly Spaced or Missing," in *Applied Time Series II*, ed. by D. F. Findley. New York: Academic Press, 1981.
[5] DUNSMUIR, W., AND P. M. ROBINSON: "Estimation of Time Series Models in the Presence of Missing Data," *Journal of the American Statistical Association*, 76 (1981), 560–568.
[6] GOURIÉROUX, CH., A. MONFORT, AND A. TROGNON: "Nonlinear Asymptotic Least Squares," paper presented at the ESEM in Pisa, Italy, 1983.
[7] HARVEY, A. C., AND P. L. V. PEREIRA: "The Estimation of Dynamic Models with Missing Observations," paper presented at the Econometric Society World Meeting in Aix-en-Provence, France, 1980.
[8] HARVEY, A. C., AND C. R. MCKENZIE: "Missing Observations in Dynamic Econometric Models," London School of Economics, Discussion Paper, 1982.
[9] KMENTA, J.: "On the Problem of Missing Measurements in the Estimation of Economic Relationships," in *Proceedings of the Econometric Society Meeting 1979*, ed. by E. G. Charatsis. Amsterdam: North-Holland Publishing Company, 1981.
[10] PALM, F. C., AND TH. E. NIJMAN: "Linear Regression Using Both Temporally Aggregated and Temporally Disaggregated Data," *Journal of Econometrics*, 19 (1982), 333–343.
[11] ———: "Consistent Estimation Using Proxy-Variables in Models with Missing Observations," paper presented at the ESEM in Pisa, Italy, 1983.
[12] ROTHENBERG, T. J.: "Identification in Parametric Models," *Econometrica*, 39 (1971), 577-591.
[13] SARGAN, J. D.: "Identification and Lack of Identification," *Econometrica*, 51 (1983), 1605–1634.
[14] TELSER, L.: "Discrete Samples and Moving Sums in Stationary Stochastic Processes," *Journal of the American Statistical Association*, 62 (1967), 484–499.
[15] WATSON, M. W., AND R. F. ENGLE: "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models," *Journal of Econometrics*, 23 (1983), 385–400.
[16] ZELLNER, A., AND C. MONTMARQUETTE: "A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analyses," *Review of Economics and Statistics*, 53 (1971), 335–342.